UNIVERSITÉ DE MONTRÉAL

CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, STATION A, MONTRÉAL, QUÉBEC, H3C 3J7

Probabilistic Networks

by

B.L. Fox

This research was supported by grants from the National Research Council of Canada and the Ford Motor Company of Canada.

Presented at the ORSA/TIMS
Joint National Meeting, Las Vegas,
Nevada, November 1975.

Université de Montréal
Publication # 217 - Département d'informatique
Publication # 33 - Centre de recherche sur les transports
December 1975

ABSTRACT

For networks with random arc lengths, we consider Monte Carlo calculation of "criticality" and path length distribution. We point out anomalies that can arise with the conventional definition of criticality and propose a new definition that to some extent gets around the problem. The conceptual and computational advantages of k-th shortest paths in this connection are indicated. Criticality determination has obvious relevance to PERT/CPM. We also exploit it for traffic assignment applications. We use novel control variates for estimating shortest (or longest) path length distribution.

We consider networks with random arc lengths. Such networks arise naturally in PERT/CPM problems and in traffic assignment. The latter application is discussed in detail in the next section. There may also be an application to so-called reliability networks.

We discuss shortest path formulations, noting that any longest path problem can be converted immediately to a shortest path problem. In applications, one wants to estimate not only the probability distribution of the length of the shortest path but also the probability that any given arc will be "critical". Roughly speaking, an arc is critical if it is on the shortest path. This concept is examined later. VAN SLYKE[11] points out the importance of criticality in the PERT context: the more critical an activity, the more likely variations in its time will affect project time and hence the more closely it should be "managed". He also indicates several worthwhile computational shortcuts for Monte Carlo estimation of criticality and path length. For the latter, BURT, GAVER, AND PERLAS[2] illustrate the importance of control variates, antithetic variates, and stratified sampling. This terminology, common in simulation, is defined in reference 2, which the reader is urged to consult. Antithetic variates, for example, can also produce significant variance reductions for criticality estimation. We recommend them.

After some preliminary sections, we consider k-th shortest paths in estimating criticalities. Later, we introduce novel control variates for estimating the distribution of shortest path length. The final section shows that this distribution may be far away from that of the path whose expected length is shortest.

TRAFFIC ASSIGNMENT

In traffic assignment, one wishes to estimate overall link flows in a multiple origin/destination context. One approach is discussed in DIAL's paper [4]. He assumes constant arc lengths and no congestion delays, but still he diverts traffic off the shortest route. One possible justification is that users perceive arc lengths incorrectly; another is that there may be secondary criteria such as cost (e.g., tolls) or scenic attraction. Weaknesses in DIAL's approach are pointed out by FLORIAN AND FOX [5], among others.

Ideally, congestion should be modelled directly. When arc lengths, given flows, are constants and origin/destination demands are also deterministic, this has been done by NGUYEN [9], for example. Even supposing that everything is deterministic, it may be difficult to estimate volume/delay curves. One would have to measure link length and link flow simultaneously over a fairly wide range of flow values. If demand is not deterministic, the predicted mean demands do not suffice when arc length depends on the amount of traffic, except as a heuristic. There is no "certainty equivalence." Acting as if there were certainty equivalence can lead to large errors. Demand forecasts are likely to be hazy, but for the sake of argument let us assume that the complete demand distributions are known. Even if it were now possible to model congestion via a mathematical program, such a program would be very

difficult to solve, probably requiring a huge Monte Carlo calculation. Whether this would be worthwhile is a subjective question.

In view of these potential difficulties, it may be preferable to model congestion via random arc lengths. One advantage is that we have to measure only variations in arc length and not flow. Another is that mean demands now suffice to estimate mean link flow. It is much easier to forecast mean demand than to estimate the complete demand distribution. Of course, it may be the demand at rush hour that is relevant. Time-dependent arc lengths and origin/destination demands can be handled in a straightforward manner.

Even if congestion delay were not a significant factor, we may anticipate randomness in arc lengths due to varying weather conditions, accidents, road repair work, etc. Random delays for traffic lights and turns are less than perfectly correlated with volume. A complication that we take account of only implicitly is that users may be risk-averse. Between two paths with approximately the same expected length, the user may choose the one with the smaller variance.

In a later section, we consider k-th shortest paths. For PERT networks, there is no possibility of path anomalies, such as paths with loops. For general networks, we impose the following (stronger) condition, also used by DIAL [4], on k-th shortest paths:

Reasonableness property: At each successive node on the path, we are farther from the origin and closer to the destination.

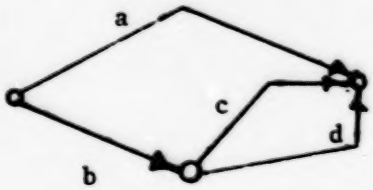
It can be argued that the reasonableness restriction is too strong: but, for computer implementation, we cannot think of a better alternative. If a k-th shortest path is generated that is not "reasonable", we would not take account of it in estimating arc criticalities.

Given these criticalities, we give a traffic assignment algorithm in the second half of the next section.

BURRELL (e.g., see POTTS AND OLIVER [10], p.65) proposes a quite different traffic assignment approach based on Monte Carlo. Each arc is assumed to have its length uniformly distributed on eight possible values evenly spread about a (link-dependent) mean value. This assumption is rather arbitrary and, we believe, unnatural. In PERT, the beta distribution has often been similarly invoked, but at least its shape can be varied depending on the choice of parameters. For each O/D, BURRELL samples the arc lengths and assigns all the traffic to the corresponding shortest route. The most damaging critique of this approach is that the estimators of overall (mean) link flows will have extremely high variances: one must add up the variances over all O/D pairs. There is no cancellation of errors. It is claimed that the evidence indicates that the assignments obtained are "realistic." This is surely accidental. The approach ignores the precept that, for statistical reliability, one must replicate. Besides replications, our approach has additional merits, as we point out.

CRITICALITY

Certain anomalies that can arise in connection with criticality are illustrated in the network below.



Here, all paths are supposed to have stochastically the same length. In particular, all expected path lengths are equal but there are small random variations. The lengths of the three paths differ only slightly. We suppose that arc b has constant length. Intuitively, all arcs are critical but not all have the same probability of being on the shortest path; arc b is twice as likely as arc a to be on the shortest path. While we would grant that arc b is perhaps more critical than arc a, we would not say that it is twice as critical. The criticality measure that we give in the next section should be taken only as indication of ordinal ranking. The numbers themselves should not be taken too seriously. For example, if we change our hypothesis on arc b and suppose that its length is random (while now the lengths of arcs c and d are constant and equal), then arcs a and b have the same chance of being on the shortest path.

In the traffic assignment context, even though all paths in the above network have the same length stochastically, we would not expect all paths to carry the same flow. We would expect an even division of flow among a and b and another even split between c and d . Thus, the path (b,c) would carry 1/4 of the flow, not 1/3. (Supposing that all arc lengths were constants, DIAL's algorithm would assign equal flows to all paths.).

Our solution to this is heuristic and not entirely satisfactory. Restrict attention to arcs whose criticality is above a given threshold, plus arcs on the path of shortest expected length. The latter are likely to be sufficiently critical anyhow, but see the last section for a counterexample. (Computations should be carried out for a range of threshold values. As a rule of thumb, we would restrict this range to the interval (.9,1).) Next, we drop nodes now disconnected from the destination and all arcs incident to them. If the reduced network contains loops, raise the threshold value. A sufficiently large threshold produces the "all-or-nothing" assignment.

Now, in the reduced network, divide incoming flow at each node evenly among outgoing arcs. This should be done recursively, starting at the origin. Since the reduced network is acyclic, no difficulties arise.

Finally, repeat the whole procedure for each origin/destination pair (O/D) and superimpose the link flows.

We would guess that the time to generate the random arc lengths dominates the time for the shortest path calculations, in line with VAN SLYKE's observations. Under this circumstance, in a single replication (i.e., with the same set of arc lengths) we would compute the (k-th) shortest routes for all O/D pairs. For a fixed number of replications, this increases the variance of the estimators of overall link flows due to positive correlations between the various shortest routes. However, this technique permits more replications for a fixed amount of computer time. We think that this is a good tradeoff.

USE OF K-TH SHORTEST PATHS

We indicate how k-th shortest paths can be used in a heuristic fashion to estimate arc criticality. For a fixed O/D, let D_{ik} be the length of the k-th shortest path in replication i ; call the corresponding set of arcs S_{ik} . In replication i, we would keep on calculating k-th shortest paths as long as $D_{ik} - D_{i1}$ stays within, say, p percent of D_{11} (for example, take $p = 5$). Suppose that m_i paths satisfy this criterion. Let $X_{ik}(a) = 1$ if a in S_{ik} , 0 otherwise. Suppose that $D_{ik} - D_{i1} = q_{ik} D_{i1}$. On replication i, the normalized "score" of arc a is

$$T_i(a) = \max [(1-q_{ik}) X_{ik}(a) : 1 \leq k \leq m_i]$$

with average score (criticality)

$$T(a) = \sum_{i=1}^{r} T_i(a)/r ,$$

where r is the number of replications. The reason for using "max" rather than "Σ" in the definition of $T_i(a)$ is an attempt to avoid anomalies in criticality values for networks like the one pictured earlier.

As an estimator of the probability that arc a is on the shortest path, T(a) is generally biased and even inconsistent. (An estimator's bias is its expected error. An estimator is consistent if its bias goes to zero as the number of observations goes to infinity.)

So we must justify T(a). As we have already indicated, the probability that arc a is on the shortest path may be a misleading measure of criticality. We think that T(a) gives a more reasonable measure. A second justification is that there are likely inaccuracies in specifying the probability distributions of arc lengths, as often there is little or no data at hand. Our measure takes account of arcs that would likely be on the shortest path if the probability distributions of arc lengths were slightly changed. In this sense, it is a robust estimator.

We now compare the (first) shortest path calculation (i) with the k-th shortest path calculation (ii). For simplicity, suppose that there are n nodes with m arcs pointing out from each node. A sharp upper bound for (i) is $O(mn)$. For (ii), $O(n(k \log m + m))$ calculations are required, as FOX [6,7] shows using HEAPSORT in conjunction with a standard algorithm. For recent discussions of HEAPSORT, see AHO, HOPCROFT, AND ULLMAN [1] and DENARDO [3]; in particular, these references show that heaps can be formed in linear

time. Comparing the computing times for (i) and (ii) we see that pays to trade first shortest path calculations for k-th shortest path calculations whenever possible. This gives a third argument for $T(a)$, with $p \neq 0$. More explicitly, there is a tradeoff between the parameters p and r above. The choice $p = 0$ gives the obvious estimator, but intuitively this would call for a rather large r to make up for the fact arcs "almost" on the shortest path would be totally ignored. We have already argued that a positive p would be reasonable even if computer time were free. The fact this economizes on the number of replications to determine a statistically reliable measure of criticality is an additional advantage.

CONTROL VARIATES

We wish to estimate the distribution of shortest path length, or perhaps merely its expectation. Control variates may lead to a variance reduction of the estimator. This is certainly not a new idea, but our control variates are novel and (to us, at least) appealing. Those familiar with "jackknifed" estimators (e.g., see GRAY AND SCHUCANY[8]) may think of the control variates Y_i below as jackknifed variates, although this (colorful) analogy is not necessary to appreciate them.

Set $d_i(a)$ equal to the length of arc a on replication i and

$$Y_i = \sum_{j \neq i} \sum_{a \in S_{j1}} [\, Ed_1(a) - d_i(a) \,] \,/\, (r-1).$$

The idea is that S_{j1} corresponds to a shortest path on replication j. We compare the expected length of this path with its actual length on replication i. We expect the latter to be positively correlated with D_{i1}. The restriction "$j \neq i$" above ensures that $EY_i = 0$. We prefer Y_i to the control variates suggested in reference 2, although we have not performed experiments to check this.

As the next section shows, it is not necessarily true that the length of the path P with shortest expected length is strongly correlated with the length of the shortest path. For this reason, we have not based a control variate on P.

Put

$$Z = \sum_{i=1}^{r} Y_i / r$$

$$D_i = D_{i1} .$$

Let $g(D_1,\ldots,D_r)$ be an estimator of some quantity of interest. For example,

$$g(D_1,\ldots,D_r) = \sum_{i=1}^{r} D_i / r$$

estimates the expected shortest path length;

$$g(D_1,\ldots,D_r) = (\text{number of } D_i\text{'s} \leq t) / (r+1)$$

estimates via the empirical distribution the probability that the length of the shortest path is $\leq t$. Instead of the estimator $g(D_1,\ldots,D_r)$, we would use the estimator $g(H_1,\ldots,H_r)$ where $H_i = D_i + Z$. If g is

linear, the new estimator is unbiased if the old one was. If g is nonlinear, the new estimator is consistent but biased. However, we would guess that its mean square error would be less. Note that Y_i and (hence) Z have expectation 0 and can be expected to be negatively correlated with D_i. To check whether using the H_i's seems worthwhile, we would divide the replications in blocks and use the empirical variances of the corresponding estimators. Carrying this idea a bit farther, we could compare jackknifed variance estimates. (Jackknifed estimates are recommended in reference 2, for example.) If r is determined sequentially to satisfy a given variance bound, our control variate trick may permit a smaller number of replications than would otherwise be possible.

As far as we know, "jackknifed" control variates have never been previously suggested. It is clear that the concept generalizes well beyond the present context. We leave a more abstract formulation to the reader.

Numerous papers seek to avoid Monte Carlo by computing bounds on the distribution of shortest path length. However, with a large number of arcs, the errors can pyramid and we would guess that examples can be constructed where the bounds are arbitrarily far off. In any case, one should generally determine arc criticalities and this requires Monte Carlo; we get a handle on the distribution of shortest path length as a virtually free byproduct.

A PATHOLOGICAL EXAMPLE

Consider a network like the one pictured earlier except that now n arcs are added in parallel to arcs c and d. We assume that the lengths of these $n+2$ parallel arcs are stochastically independent and distributed uniformly on the interval $[100,200]$. Let the length of a (b) be 240 (100). Thus, path a has shortest expected length. But, by taking n sufficiently large, we make the expected length of the shortest path arbitrarily close to 200. The probability that arc a (b) [c] is on the shortest path is $.6^{n+2}$, $(1-.6^{n+2})$, $[(1-.6^{n+2})/(n+2)]$. So arc a, the path with the shortest expected length, has the smallest chance, by an order of magnitude, of being on the actual shortest path. This shows that the "certainty equivalence" approach can lead to arbitrarily large errors.

If our traffic assignment algorithm were applied to this network, there would be a discontinuity depending on the criticality threshold. Arc a carries all the flow, unless the criticality threshold is set absurdly low. This assignment is intuitively reasonable.

REFERENCES

1. A.V. AHO, J.E. HOPCROFT, AND J.D. ULLMAN, The Design and Analysis of Computer Algorithms, Addison-Wesley, Reading, Mass. 1974.

2. J.M. BURT, Jr., D.P. GAVER, AND M. PERLAS, "Simple Stochastic Networks: Some Problems and Procedures," Naval Res. Logistics Quarterly 17, 439-460 (1970).

3. E.V. DENARDO, Dynamic Programming: Theory and Application, to be published.

4. R.B. DIAL, "A Probabilistic Multipath Assignment Model which Obviates Path Enumeration," Transportation Research 5, 83-111 (1971).

5. M. FLORIAN AND B. FOX, "On the Probabilistic Origin of Dial's Multipath Traffic Assignment Model," Publication # 209, Département d'Informatique, Université de Montréal, 1975.

6. B.L. FOX, "Calculating k-th Shortest Paths," INFOR 11, 66-70 (1973).

7. B.L. FOX, "More on k-th Shortest Paths," Comm. ACM 18, 279 (1975).

8. H.L. GRAY AND W.R. SCHUCANY, The Generalized Jackknife Statistic, Marcel Dekker, Inc., New York 1972.

9. S. NGUYEN, "An Algorithm for the Traffic Assignment Problem," Transportation Science 8, 203-216 (1974).

10. R.B. POTTS and R.M. OLIVER, Flows in Transportation Networks, Academic Press, New York 1972.

11. R.M. VAN SLYKE, "Monte Carlo Methods and the PERT Problem," Operations Research 11, 839-860 (1963).